DOCUMENT RESUME

ED 331 890 TM 016 497

AUTHOR Fluegge, Lynn
TITLE The Evaluation of Intern Teachers: The Development of an Observation Instrument.
INSTITUTION Kentucky State Dept. of Education, Frankfort. Office of Research and Planning.
PUB DATE Mar 90
NOTE 34p.
PUB TYPE Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Classroom Observation Techniques; *College Faculty; *Elementary School Teachers; Elementary Secondary Education; Higher Education; *Secondary School Teachers; Teacher Education; Teacher Evaluation; *Teacher Interns; *Test Construction; Test Reliability; Test Validity; Videotape Recordings
IDENTIFIERS *Kentucky Teacher Internship Program; Test Retest Reliability

ABSTRACT

The validation of a classroom observation instrument to be used in evaluating teacher interns under the Kentucky Teacher Intern Program is described. The validation project began in 1988. Fourteen educators from state universities comprised a jury that judged the initial version of the instrument and compared it with the instrument previously used, an observation instrument from the Florida Performance Measurement System. Participants applied the instrument to videotaped samples of classroom instruction. For determination of test-retest reliability, subjects from the validation phase oriented 28 local educators with regard to the use of the revised instrument. Each participant coded two videotapes and one actual intern in the classroom. This test established that the instrument had sufficient reliability and validity to be used in the Kentucky program. Data indicate the need for a small group of knowledgeable people to revise the coding manual and make some revisions to the coding instrument. Thirteen tables are included. Five appendices present the original and revised Kentucky Teacher Internship Program Classroom Observation Instruments and supplemental information about the development of the instruments. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED331890

THE EVALUATION OF INTERN TEACHERS:

The Development of an Observation Instrument

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

March, 1990

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

LYNN FLUEGGE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

Lynn Fluegge, Ph.D.
Kentucky Department of Education
Office of Research and Planning
Division of Research

TM016497

Contents

Introduction 1

Project Rationale 2

Instrument Development 2

Pilot Test 8

Data Analysis 9

Table 1A 3
Table 1B 4
Table 1C 4
Table 1D 5
Table 1E 5
Table 1F 6
Table 1G 6
Table 2 7
Table 3 8
Table 4 10
Table 5 11
Table 6 12
Table 7 13

Appendix A A-1
Kentucky Beginning Teacher Internship Program Classroom Observation Instrument

Appendix B B-1
Summary of Participant Data for the August, 1988 Pilot Phase of the Teacher Intern Program

Appendix C C-1
Kentucky Beginning Teacher Internship Program Pilot Project

Appendix D D-1
Kentucky Beginning Teacher Internship Program Pilot Research Project Questionnaire

Appendix E E-1
Kentucky Beginning Teacher Internship Program Classroom Observation Instrument

THE EVALUATION OF INTERN TEACHERS:
The Development of an Observation Instrument

INTRODUCTION

When it began, the Kentucky Teacher Intern Program (KTIP) used materials developed as part of the Florida Performance Measurement System (FPMS); however, modifications prompted the Kentucky Department of Education to develop its own materials to use in observing the classroom behaviors of Kentucky's beginning teachers. The evolution of a specific Kentucky instrument began in 1984 when the General Assembly enacted legislation which established the Kentucky Career Ladder Commission. That Commission did a considerable amount of developmental work in constructing a classroom observation instrument on which student behaviors could be coded and summarized.

In its final report, the Commission documents the steps involved in the development of a coding instrument and coding manual to determine if a standard format could be used to observe a classroom teacher and later used to determine if the teacher should receive a rating of satisfactory or outstanding. The classroom observation instrument was one part of a multipronged process used to determine if the teacher would receive a monetary reward for outstanding work.

In the process of developing the observation instrument, the Career Ladder Commission reviewed national efforts to help and support first year teachers. The process included reviewing various induction models being used nationwide. The FPMS coding instrument appears in Peterson, Kromrey, Micceri, and Smith (1987), and the Kentucky Career Ladder instrument appears in the Commission's Final Report (1988).

The Advisory Committee for Teacher Improvement (ACTIP), an advisory group attached to the Division of Teacher Education and Certification, had to find a new instrument for collecting classroom data and wanted to begin where the Career Ladder Commission ended. ACTIP asked the Division of Research to determine the reliability and validity of this instrument (Appendix A).

The entire validation project began in the spring of 1988, and the instrument was to be ready for field use by May, 1989. By inference, items could be removed from the KTIP because of data resulting from the project, but no items could be added because data would not be available to justify inclusion of untested items. The product also had to be more 'user friendly' than the FPMS classroom observation instrument.

PROJECT RATIONALE

The Division of Research developed the research methodology, collected data, and analyzed both. The Division of Teacher Education and Certification scheduled regional meetings and paid participant expenses. The initial research project was to be completed to allow the instrument to be used in the elementary and secondary schools during the 1989-90 school year. The research project had to accomplish the following major tasks:

1. Establish validity for the Kentucky instrument.
2. Determine the instrument's utility and its raters' ability to use it accurately and consistently.
3. Collect data which would aid in producing a revised Kentucky instrument and coding manual.

The project staff utilized two pre-experimental research designs (Stanley and Campbell 1963) to accomplish the three tasks. The entire project would create four separate data bases while using six training video tapes - two elementary, two middle school, and two high school tapes - reviewed by various subjects at different times and coded using the FPMS or/and KTIP instrument.

The content validity of the newly developed KTIP classroom observation instrument was effectuated by employing a process similar to that expressed by Kerlinger (1987). The ACTIP had to develop an instrument that would be easier for observers to code than the FPMS, yet its form and content had to be more accessible and adequately represent observable teacher behaviors.

Determining the reliability of the KTIP instrument required anchoring it to another instrument that had some record of reliability and validity, and the FPMS was the obvious choice. A jury of experts inspected the KTIP instrument and determined which items therein were exact or partial representations of items on the FPMS instrument. Those six people represented staff members from the Divisions of Research and Teacher Education and Certification and University staff members who had been involved in the Career Ladder Commission work and the teacher intern program. The crosswalk emphasized those teacher behaviors which were coded by making hash marks. Behaviors reflected in items on the back of the KTIP and on the cover of the FPMS document indicated with a check mark were not included in this analysis. The six crosswalkers represented the spectrum of potential jurists. No overt selection process restricted the variation of the six raters.

INSTRUMENT DEVELOPMENT

Tables 1A through 1G array the jury of experts' consensus markings as determined by the Division of Research staff. KTIP category refers to subparts of seven sequential sections inside the coding instrument from Content 1, "States/

defines content focus" through Time 1 "Mismanages instructional time". The FPMS category indicates items from that instrument which crosswalkers matched to the KTIP category listed on the left. For example, crosswalker #1 matched 3L "Orients students to classwork/maintains academic focus" on FPMS to Content 1 "States/defines content focus". Crosswalker #2 identified 11L "Treats concept-definition/attributes/examples/non-examples" as well as 13L "States and applies academic rule" as partial or complete matches. Since FPMS utilizes left (L) and right (R) markings, an algorithm was built which utilized certain L behaviors in some cases and some R behaviors in other cases.

Table 1A. displays the jury's findings in the content section. This group matched all seven of the content teaching behaviors with items on the FPMS.

TABLE 1A.
Cross-Reference Matrix
KTIP Classroom Observation Instrument (Content Section)
Florida Performance Measurement System

KTIP Category		FPMS Category	
States/defines content focus	(C1)	Orients students to classwork/ maintains academic focus	(3L)
		Treats concept-definition/attributes/ examples/nonexamples	(11T)
Explains content clearly	(C2)	Treats concept-definition/attributes examples/non-examples	(5AR)
		States and applies academic rule	(13L)
Models/applies content focus	(C3)	Treats concept-definition/attributes examples/non-examples	(11L)
		Discusses cause-effect/uses linking words/applies law or principle	(12L)
Checks for comprehension	(C4)	Gives directions/assigns/checks comprehension of homework, seatwork assignment/gives feedback	(9L)
		Questions: academic comprehension/ lesson development	(5A,BL)
Provides guided practice	(C5)	Provides for practice	(BL)
		Circulates and assists students	(10L)
Conducts review/summary	(C6)	Conducts beginning/ending review	(4L)
Provides independent practice	(C7)	Gives directions/assigns/checks comprehension of homework, seatwork assignment/gives feedback	(9L)

Table 1B. deals with the use of questioning techniques. Of the nine items on the KTIP, the experts found five matches.

TABLE 1B.

Cross-Reference Matrix
KTIP Classroom Observation Instrument (Questioning Techniques Section)
Florida Performance Measurement System

KTIP Category		FPMS Category	
Asks academic question	(QT1)	Questions: academic comprehension/ lesson development	(5A,BL)
Asks multiple question	(QT2)	Poses multiple questions asked as one,unison response	(5AR)
Asks/pauses/names	(QT3)		
Does not provide wait	(QT4)		
Guides reciter	(QT5)		
Does not guide reciter	(QT6)		
Allows call-outs	(QT7)	Poses multiple questions asked as one, unison response	(5AR)
Asks procedural questions	(QT8)	Poses non-academic questions/ nonacademic procedural questions	(5BR)
Asks unrelated questions	(QT9)	Poses non-academic questions/ nonacademic procedural questions	(5BR)

Table 1C. analyzes the section detailing the intern's reactions to student responses. In three of six response situations, the jury found similarity.

TABLE 1C.

Cross-Reference Matrix
KTIP Classroom Observation Instrument (Responses Section)
Florida Performance Measurement System

KTIP Category		FPMS Category	
Acknowledges student responses	(R1)	Recognizes response/amplifies/gives corrective feedback	(6L)
Rephrases/amplifies student responses	(R2)	Recognizes response/amplifies/gives corrective feedback	(6L)
Corrects/clarifies student responses	(R3)	Recognizes response/amplifies/gives corrective feedback	(6L)
Does not correct/clarify responses	(R4)		
Responds to academic questions/input	(R5)		
Responds to non-academic questions/input	(R6)		

Table 1D. depicts the manner in which both instruments allow the reviewer to record praise. The group established matches in three of four praise areas.

TABLE 1D.

Cross-Reference Matrix
KTIP Classroom Observation Instrument (Praise Section)
Florida Performance Measurement System

KTIP Category		FPMS Category	
Uses specific academic praise	(P1)	Gives specific academic praise	(7L)
Uses general academic praise	(P2)	Uses general, non-specific praise	(7R)
Uses group academic praise	(P3)	Uses general, non-specific praise	(7R)
Uses conduct-related praise	(P4)	_____	

Table 1E. concerns student motivation and is labeled communication. The cross walkers discerned parallels two of four times.

TABLE 1E.

Cross-Reference Matrix
KTIP Classroom Observation Instrument (Communication Section)
Florida Performance Measurement System

KTIP Category		FPMS Category	
Cues students	(CO1)	_____	
Uses emphasis	(CO2)	_____	
Uses challenge	(CO3)	Expresses enthusiasm/verbally challenges students	(16L)
Uses sarcasm/negative effect	(CO4)	Ignores student or response/expresses sarcasm, disgust, harshness	(6R)

Table 1F. compares methods of dealing with student misconduct. The jury detected comparisons for all four KTIP items.

TABLE 1F.

Cross-Reference Matrix
KTIP Classroom Observation Instrument (Management Section)
Florida Performance Measurement System

KTIP Category		FPMS Category	
Stops misconduct positively	(M1)	Stops misconduct	(20L)
Stops misconduct negatively	(M2)	Delays desist/doesn't stop misconduct/desists punitively	(20R)
Does not stop misconduct	(M3)	Delays desist/doesn't stop misconduct/desists punitively	(20R)
Uses sarcasm/negative effect	(M4)	Maintains instructional momentum	(21L)

Table 1G. shows how the two instruments deal with appropriate use of time in the classroom. Again both items had numerous matches on the FPMS.

TABLE 1G.

Cross-Reference Matrix
KTIP Classroom Observation Instrument (Time Section)
Florida Performance Measurement System

KTIP Category		FPMS Category	
Minimizes management time	(T1)	Begins instruction promptly	(1L)
		Handles materials in an orderly manner	(2L)
		Maintains instructional momentum	(21L)
Manages instructional time	(T2)	Delays	(1R)
		Does not organize or handle materials systematically	(2R)
		Loses momentum - fragments non-academic directions, overdwells	(21R)

A review of all seven subsections of Table 1. establishes a consensus from the jury of experts that 26 of the 36 KTIP teaching behaviors are recordable on the

FPMS. Both forms contained extraneous items such as the physical features of the classroom and years of teaching experience. Because many of the matching items were like these and not considered, the similarity depicted between the two instruments in Table 1. is understated.

During the period of time from May 31, 1988 through June 3, 1988 two representatives from each of the eight state universities met in Bowling Green, Kentucky to participate in process development and data collection. Fourteen of the 16 attendees made up the participant pool. On the first day the participants viewed the six tapes from the Florida system and coded on FPMS instruments in accordance with FPMS standards and practices. Through the next three days the fourteen participants were trained to use the KTIP instrument and discussed detailed differences between FPMS and KTIP instruments. Prior to closing the session on Friday, each participant again viewed the same six FPMS video tapes. This time the participants marked the behaviors viewed on the KTIP instrument. The data gleaned from these codings became two of the four data bases.

The Division of Research examined the degree of correlation between expert responses on the 26 common items and found it to be significant. Summarized in TABLE 2, the analysis compares both codings on the six video tapes by each of the 14 subjects. Each subject had a total of 156 paired scores. Those paired scores were used to calculate the correlation coefficient for each subject. Based on video tapes which were samples of instruction, the range of markings was rather narrow. The correlation coefficients were significant beyond the .001 level for each of the fourteen subjects. When the correlations were examined by video tape for each subject, 79 of the 84 analyses resulted in statistically significant correlations at the .05 level.

TABLE 2.

Concurrent Validity (Pearson Product-Moment Correlation Coefficients)
KTIP and FPMS ratings
Using Table 1A-1G Cross-Reference

Subject Number	r for Six Tapes Combines
1045	0.6198
1264	0.6191
1380	0.5858
2581	0.5854
3527	0.5786
4560	0.6739
5101	0.5340
5366	0.4670
6201	0.6133
6268	0.6332
6788	0.5892
7188	0.6550
8810	0.6784
9333	0.5349

$p < .001$ for all correlations

Table 3. summarizes the individual correlation coefficients for each subject for each video tape. Only five of 84 comparisons produced nonsignificant results.

TABLE 3.

Concurrent Validity (Pearson Product-Moment Correlation Coefficients)
KTIP and FPMS ratings
Using Table 1A-1G Cross-reference

	TAPE 1	TAPE 2	TAPE 3	TAPE 4	TAPE 5	TAPE 6
1045	0.76	0.74	0.65	0.47	0.70	0.70
1264	0.66	0.71	0.72	0.52	0.61	0.74
1380	0.56	0.71	0.75	0.55	0.62	0.68
2581	0.54	0.71	0.70	0.48	0.50	0.70
3527	0.59	0.62	0.51	0.46	0.34*	0.70
4560	0.49	0.71	0.72	0.35*	0.68	0.76
5101	0.54	0.65	0.54	0.39	0.33*	0.59
5366	0.48	0.56	0.71	0.14*	0.67	0.67
6201	0.41	0.69	0.49	0.48	0.49	0.62
6268	0.69	0.68	0.55	0.55	0.68	0.68
6788	0.66	0.71	0.62	0.46	0.58	0.68
7188	0.69	0.74	0.69	0.54	0.58	0.70
8810	0.55	0.73	0.71	0.59	0.68	0.70
9333	0.54	0.71	0.58	0.32*	0.57	0.65

** nonsignificant result*

PILOT TEST

The second task, determination of test-retest reliability, used the subjects from the validation phase to orient other subjects in the use of the KTIP classroom observation instrument. No control or comparison group existed to contrast with the oriented groups. The design called for the trainers representing the state universities to select from their service area 28 local educators to receive two day orientation sessions. Each of the seven regional groups was to contain two participants from the following classifications. The result would be 28 participants in each region: elementary principal, middle school principal, high school principal, vocational educators, college elementary specialists, college middle school specialists, college high school specialists, elementary resource teachers, middle school resource teachers, high school resource teachers, elementary instructional supervisors, high school instructional supervisors, exceptional educators (elementary, secondary or college).

Persons besides those listed were welcome to attend the orientation sessions. Not all sessions could include all of the 28 people specified. Each parti-

cipant completing all phases received a stipend for services rendered. Orientation sessions were conducted in seven regions; as a cost saving measure one region did not participate in the reliability determination phase. Appendix C tables group characteristics. Each participant had to code two tapes as part of the orientation. Each person then had ninety days to code an actual intern in the classroom using the KTIP instrument. The participants were also instructed to make any written comments they deemed appropriate about the instrument or coding manual. The comments would be turned in when the participants returned to the orientation site again to code two tapes, which would become the retest data set.

The seven orientation sessions were held between late July, 1988 and August, 1988. Each orientation session had the same agenda and each was conducted by local staff members. The morning portion called for reviewing the KTIP coding manual, coding practice video tapes, and other activities to get the local educators acclimated to the KTIP Instrument. In the afternoon portion, the group was divided by level (elementary, middle, or secondary). Each group viewed two of the six video tapes designated for its level and used the KTIP instrument to code. This design allowed the data to be sorted by geographical area, by level, and by responsibility and served as a pretest to determine test-retest reliability while building data base three.

DATA ANALYSIS

The instrument yielded three different data subsets for each category. The three subsets were: (1) content determination, (2) hashmark section, and (3) checkmark section. The content area presented the largest problem in quantification because each coding participant saw different content being taught. Some people viewed a tape and saw many different content areas; others saw a single content being taught. Later discussions revealed that these differences resulted from the background of participants in specific curriculum areas. Subjects who were knowledgeable in a certain area tended to indicate a larger number of components than an individual who was not as highly versed in that specialty area. The participant was required to list the content in the slant lines above the content descriptors and place a check mark after each of the seven content items if/as the teacher on the video tape completed each. The data set resulted in a variety of marking patterns for the same tape. One should also note that the concept of 'content' checkmarks and writing the content area was new to some of the participants. Although a similar procedure was used in the Career Ladder project, only a small sample of the participants in this project had used the procedure.

Since the sessions were for orientation purposes, the participants did not get a great amount of time to practice coding this section. One participant could have marked one content area while another marked four content areas for the same tape. Table 4 arrays a summary of pre- and post markings. With many possible markings, a rigorous statistical analysis was not appropriate. The table entries report the resultant distribution of checkmarks as being either unimodal or multimodal. In the

pretest distributions, tapes 1, 5, and 6 were marked as either one or two modes, not mixed, while tapes 2, 3, and 4 were mixed. The number of subjects marking each tape is reported beside the "N". A large difference in the participant count resulted because one orientation site did not collect data in the same design as the other six sites. This inconsistency caused the total participants responding to tapes 3 and 4 to be 47 and 42 on the pretest. Eight reversals in modality occurred from the pretest to the posttest which could have been the result of numerous intervening events such as: field practice; question and answer sessions held before the posttest data was collected; or exchanges between various participants as part of working as members of an intern committee. No effort was made to keep participants from communicating or doing other educational activities which might influence the marking patterns. Consistency existed from the pretest to the posttest on tapes 1, 5, and 6; however, one reversal was noted on tape 4; two reversals on tape 2; and five reversals on tape 4. Overall, 80.9% or 34 out of 42 categories retained their original modality. Tape 2 was an elementary tape and tapes 3 and 4 were middle school tapes.

A chi square analysis performed on the two data matrices used the pre-matrix as the expected outcome for the post matrix. In the prematrix 27 cells were one mode and 15 cells were bimodal cells resulting in a chi square value of .30. No significant difference occurred when comparing the cells of the posttest to the cells of the pretest.

TABLE 4.

Cross-Reference Matrix
Modality of Markings, Pre- and Posttest, Six Tapes
Seven Content Areas

		TAPE 1		TAPE 2		TAPE 3		TAPE 4		TAPE 5		TAPE 6	
		Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
States/defines content focus	(C1)	ONE	ONE	ONE	TWO	ONE	ONE	ONE	ONE	TWO	TWO	ONE	ONE
Explains content clearly	(C2)	ONE	ONE	TWO	TWO	ONE	TWO	ONE	ONE	TWO	TWO	ONE	ONE
Models/applies content focus	(C3)	ONE	ONE	TWO	TWO	ONE	TWO	TWO	TWO	TWO	TWO	ONE	ONE
Checks for comprehension	(C4)	ONE	ONE	TWO	TWO	TWO	ONE	ONE	ONE	TWO	TWO	ONE	ONE
Provides guided practice	(C5)	ONE	ONE	ONE	TWO	ONE	TWO	TWO	ONE	TWO	TWO	ONE	ONE
Conducts review/summary	(C6)	ONE	ONE	TWO	TWO	ONE	ONE	ONE	ONE	TWO	TWO	ONE	ONE
Provides independent practice	(C7)	ONE	ONE	ONE	ONE	TWO	ONE	ONE	ONE	TWO	TWO	ONE	ONE
	N =	70	66	69	66	47	41	42	38	67	64	67	64

Table 5. depicts the hashmark sections from Questioning Technique 1 through Time 2. The table represents the number of times subjects marked the categories while viewing the six tapes during both marking sessions. The "TOT IN" column shows the number of subjects who marked within the 70% range established

by the fourteen experts during the first meeting. The data indicate that 72.29% of the 20,329 grids marked fell within acceptable standards with the percentages being slightly higher for pretest (75.27%) than for posttest (73.25%). The division of research made no effort to validate additionally the criteria established during the May-June meeting. The pretest/posttest correlation was .97 for the items.

TABLE 5.

Number and Percentage of Subjects Marking Within the Criterion of 70 Percent by Category for All Six Tapes Pre- and Post-test

CATEGORY		TOT IN	TOT OUT	TOT TOT	% IN
Asks academic question	(QT1)	472	229	701	67.33
Asks multiple questions	(QT2)	405	296	701	57.77
Asks/pauses/names	(QT3)	406	295	701	57.92
Does not provide wait time	(QT4)	623	78	701	88.87
Guides reciter	(QT5)	584	117	701	83.31
Does not guide reciter	(QT6)	701	0	701	100.00
Allows call-outs	(QT7)	400	301	701	57.06
Asks procedural question	(QT8)	368	333	701	52.50
Asks unrelated question	(QT9)	623	78	701	88.87
Acknowledges student responses	(R1)	293	408	701	41.80
Rephrases/amplifies student responses	(R2)	378	323	701	53.92
Corrects/clarifies student responses	(R3)	538	163	701	76.75
Does not correct/clarify responses	(R4)	694	7	701	99.00
Responds to academic questions/input	(R5)	535	166	701	76.32
Responds to non-academic questions/input	(R6)	558	143	701	79.60
Uses specific academic praise	(P1)	535	166	701	76.32
Uses general academic praise	(P2)	423	278	701	60.34
Uses group academic praise	(P3)	650	51	701	92.72
Uses conduct-related praise	(P4)	631	70	701	90.01
Cues students	(CO1)	167	534	701	23.82
Uses emphasis	(CO2)	501	200	701	71.47
Uses challenge	(CO3)	634	67	701	90.44
Uses sarcasm/negative affect	(CO4)	634	67	701	90.44
Stops conduct positively	(M1)	554	147	701	79.03
Stops conduct negatively	(M2)	643	58	701	91.73
Does not stop misconduct	(M3)	656	45	701	93.58
Manages overlapping events	(M4)	612	89	701	87.30
Minimizes management time	(T1)	358	343	701	51.07
Mismanages instructional time	(T2)	527	174	701	75.18
		15,103	**5,226**	**20,329**	**74.29**

Table 6. reports the percentage of participants marking within the 70% range when broken down by tape pre and post. Note that the percent correct decreased in five out of six tapes from pretest to posttest. Tape 5 had 73.1% correct in the pretest and improved one percent to 74.1% in the posttest. The results from pre to posttest resulted in a loss ranging from .9% on tape 6 to 4.3% on tape 3. Ordinarily, the retest percentage correct would tend to decrease slightly over time as was the case in five of the six tapes. The markings remained stable over the time between pre and posttesting.

TABLE 6.

Number and Percentage Correct for Pre- and Post-test by Tape
Hashmark Section

	Pre				Post			
Tape Number	TOT IN	TOT OUT	TOT TOT	%	TOT IN	TOT OUT	TOT TOT	%
1	1,360	670	2,030	67.0	1,124	690	1,914	63.9
2	1,576	425	2,001	78.8	1,463	451	1,914	76.4
3	1,078	285	1,363	79.1	889	300	1,189	74.8
4	879	339	1,218	72.2	749	353	1,102	68.0
5	1,421	522	1,943	73.1	1,376	480	1,856	74.1
6	1,588	355	1,943	81.7	1,500	356	1,856	80.8

Table 7. presents the number and percent of subjects marking the checklist on the back of the KTIP classroom observation instrument. To be included in the "WITHIN RANGE" group, eight of the eleven items had to have been marked correctly. This criteria resulted in 89.23% of the markings being within the range on the pretest and 84.12% within range on the posttest. The least correctly marked tape on both the pretest and posttest was Tape 4. The resultant percentages indicate that tapes 1 through 5 had a higher percentage correct on the pretest than on the posttest. The largest loss was on Tape 5 which eroded 14.62 percent. Tape 6 markings resulted in a higher percent correct on the posttest than on the pretest. Overall some net reduction in correctness was expected with a 90 day interval between the two data collections. The chi squared analysis comparing the proportions of respondents by in or out of range categories with the expected ranges from the pretest resulted in a value of 6.43 which is not significant at the .20 level.

TABLE 7.

Number and Percent of Raters Meeting the Criteria for Teaching Behaviors (Checklist on Back of Coding Form)

	Pre				Post			
Tape Number	Within Range	Out of Range	Total Marking	Percentage Accuracy in Range	Within Range	Out of Range	Total Marking	Percentage Accuracy in Range
1	61	9	70	87.14	54	12	66	81.82
2	66	3	69	95.65	61	5	66	92.42
3	43	4	47	91.49	38	4	42	90.48
4	33	9	42	78.57	25	13	38	65.79
5	59	8	67	88.06	47	17	64	73.44
6	61	6	67	91.04	61	3	64	95.13
Total	**323**	**39**	**362**	**89.23**	**286**	**54**	**340**	**84.12**

In an effort to aid in producing a revised instrument and coding manual, the participants agreed to record comments and suggestions. The comments followed a consistent pattern from site to site, which was remarkable, because the presenters used various formats to present the orientation of the materials and the response. Some allowed participants to record comments in no particular format; others used a form developed by Kentucky State University (Appendix C) or the form developed by Murray State University (Appendix D).

The comments most frequently made dealt with the location of a specific piece of information. The next most referenced commentary involved the suggested inclusion of another category to mark or exclusion of one or more of the current categories. The items which were new, such as the content areas and the coding manual definition of how to mark, received some commentary, along with the inconsistent marking patterns. A few people commented on a philosophical point or basis which was previously discussed but not documented thoroughly in the coding manual.

The pre-experimental design established that the instrument had sufficient reliability and validity to be used in the first year of the Kentucky program. The data collected indicated the need for a very small group of knowledgeable people to review comments and concerns, to revise the coding manual, and to perform minor revisions on the coding instrument. These suggested revisions could still be accomplished within the timeline of the overall process. Appendix E is the final copy of the observation instrument.

Bibliography

Kentucky Career Ladder Commission Summary Report on the 1986-87 Pilot Program. (1987). Frankfort, Ky: Kentucky Career Ladder Commission.

Peterson, D., Kromrey, J., Micceri, T., & Smith, B. O. (1987). Florida Performance Measurement System: An Example of its Application. The Journal of Educational Research, 80(3), 141-148.

Downie, N. M. and Heath, R. W. (1970). Basic Statistical Methods. (3rd ed.). New York:Harper & Row, Publishing.

Campbell, Donald T. and Stanley, Julian C. (1963). Experimental and Quasi-Experimental Designs for Research. Chicago:Rand McNally College Publishing Company.

Kerlinger, Fred N. (1986). Foundations of Behavioral Research. (3rd ed.). New York:Holt, Rinehart and Winston, Inc.

APPENDIX A

Kentucky Beginning Teacher Internship Program
Classroom Observation Instrument

Intern's Name (Print) _____ Last _____ First _____ Middle

Intern's SS# _____ Date of Observation _____

School Name _____ School Number _____

District Name _____ District Number _____

Starting Time _____ Ending Time _____

1. Check method(s) used in observed lesson.

___1. Lecture
___2. Discussion, Recitation, Interaction
___3. Independent Activity
___ 4. Combination of 1 and 2
___ 5. Combination of 1 and 3
___ 6. Combination of 2 and 3
___ 7. Combination of all

2. Check the subject area observed.

___1. Language Arts/English
___2. Mathematics
___3. Science
___4. Social Studies
___5. Music
___6. Art
___ 7. Physical Education
___ 8. Industrial Education (5-12)
___ 9. Home Economics Education (5-12)
___10. Vocational Education _____
___11. Special Education _____
___12. Other _____

3. Check type of classroom in which observation occurred.

___1. Self-contained Classroom
___2. Field or Court
___3. Library/Media Center
___ 4. Laboratory/Shop
___ 5. Resource Room
___ 6. Other _____

Provide the following information for the class observed.

4. _____ Grade level(s) of students
5. _____ Total number of students in class
6. _____ Number of students having a IEP
7. _____ Intern's number of completed years teaching experience

Length of coding (in minutes) _____

Observer's name (print) _____

Observer's signature _____

Intern's signature _____

LESSON COMPONENTS

TEACHING BEHAVIORS

Content		1. states/defines content focus																
		2. explains content clearly																
		3. models/applies content focus																
		4. checks for comprehension																
		5. provides guided practice																
		6. conducts review/summary																
		7. provides independent practice																

				SEATWORK	
Questioning Techniques		1. asks academic questions			
		2. asks multiple questions			
		3. asks/pauses/names			
		4. does not provide wait time			
		5. guides reciter			
		6. does not guide reciter			
		7. allows call-outs			
		8. asks procedural questions			
		9. asks unrelated questions			
Responses		1. acknowledges student responses			
		2. rephrases/amplifies student responses			
		3. corrects/clarifies student responses			
		4. does not correct/clarify responses			
		5. responds to academic questions/input			
		6. responds to nonacademic questions/input			
Praise		1. uses specific academic praise			
		2. uses general academic praise			
		3. uses group academic praise			
		4. uses conduct-related praise			
Communication		1. cues students			
		2. uses emphasis			
		3. uses challenge			
		4. uses sarcasm/negative affect			
Management		1. stops misconduct positively			
		2. stops misconduct negatively			
		3. does not stop misconduct			
		4. manages overlapping events			
Time		1. minimizes management time			
		2. mismanages instructional time			

OBSERVATION NOTES:

DURING THIS OBSERVATION, THE INTERN:

1. _____ began instruction promptly.
2. _____ secured the attention of students.
3. _____ reviewed students' previous work or previous lesson content.
4. _____ provided an overview of lesson content or a statement of lesson objective(s).
5. _____ provided an overview of how the lesson would proceed.
6. _____ elicited responses from a variety of students in a nonrepetitive pattern or used ordered turns (only in K-3 math & reading).
7. _____ used appropriate vocabulary/correct grammar.
8. _____ provided clear direction for seatwork/homework (write NA if not applicable).
9. _____ circulated and monitored students' progress during seatwork (write NA if not applicable).
10. _____ was involved.
11. _____ was enthusiastic.

KDE/MIC APP
3210-1037

APPENDIX B

SUMMARY OF PARTICIPANT DATA FOR THE AUGUST, 1988
PILOT PHASE OF THE TEACHER INTERN PROJECT

TAPE CODED

	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	69	19.1	69	19.1
2	68	18.8	137	37.8
3	47	13.0	184	50.8
4	42	11.6	226	62.4
5	68	18.8	294	81.2
6	68	18.8	362	100.0

ASSIGNED TEACHING LEVEL

	Frequency	Percent	Cumulative Frequency	Cumulative Percent
ELEMENTARY	139	38.4	139	38.4
MIDDLE	89	24.6	228	63.0
HIGH	134	37.0	362	100.0

PILOT SITE ATTENDED

	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NORTHERN KY	52	14.4	52	14.4
EASTERN KY	36	9.9	88	24.3
MOREHEAD ST	56	15.5	144	39.8
KY STATE	58	16.0	202	55.8
U OF L	50	13.8	252	69.6
MURRAY ST	54	14.9	306	84.5
U OF K	56	15.5	362	100.0

OCCUPATION OF PARTICIPANT

	Frequency	Percent	Cumulative Frequency	Cumulative Percent
PRINCIPAL	79	21.8	79	21.8
RESOURCE	85	23.5	164	45.3
SUPERVISOR	80	22.1	244	67.4
COLLEGE	82	22.7	326	90.1
EXCEPT ED	14	3.9	340	93.9
VOC ED	22	6.1	362	100.0

APPENDIX C

KENTUCKY BEGINNING TEACHER INTERNSHIP PROGRAM
PILOT PROJECT

Clifton L. McMahon, Ph.D., Regional Coordinator

EVALUATION OF CLASSROOM OBSERVATION INSTRUMENT

This form is to be completed immediately after the observation and brought to the meeting on November 19.

I. General Information Section:

1. Does the form call for all needed information?

2. Is unnecessary information requested?

3. Are all the items and categories clear and adequate?

4. Other comments?

II. Teaching Behaviors:

1. Are the seven areas of behavior adequate?

2. What difficulties were encountered in coding and tallying the intern's performance?

(CONTINUED ON BACK)

3. Should adjustments (additions-deletions) be made in the items in any of the seven areas?

4. Was the section for tallying seat work useful?

5. Other comments?

III. Post-Observation Check Section:

1. Did you encounter problems in recalling behavior in order to complete this section?

2. Is there an advantage to checking this section after the conclusion of the lesson?

3. What adjustments should be made to this section?

4. Other comments?

APPENDIX D

Kentucky Beginning Teacher Internship Program
PILOT RESEARCH PROJECT QUESTIONNAIRE
November 16, 1988

1. How well prepared were you (training-wise) to code this instrument in the field?

2. How difficult do you think it will be to train the teachers and administrators in our service region to reliably code with this instrument?

3. Specifically... what is the minimum number of days which will be required to train administrators and teachers in our region to reliably code with this instrument?

4. Is this instrument easier or harder to use in the field than the FPMS instrument?

5. Specifically... in what ways is this instrument easier to use than the FPMS?

6. Specifically... in what ways is this instrument harder to use than the FPMS?

7. What recommendations do you make for the use of the subject matter section which will make it easier/better to use.

8. What do you think will be the attitude of your colleagues towards this instrument once they have been trained with it and start to use it?

APPENDIX E

Kentucky Teacher Internship Program
Classroom Observation Instrument

Intern's Name _____
(Print) Last First Middle

Intern's SS# _____ Date of Observation _____

School Name _____ School Number _____

District Name _____ District Number _____

Starting Time _____ Ending Time _____

1. Check the subject area observed.

___1. Language Arts/English
___2. Mathematics
___3. Science
___4. Social Studies
___5. Music
___6. Art
___7. Physical Education
___8. Industrial Education (5-12)
___9. Home Economics Education (5-12)
___10. Vocational Education _____
___11. Special Education _____
___12. Other _____

2. Check type of classroom in which observation occurred.

___1. Self-contained Classroom
___2. Field or Court
___3. Library/Media Center
___4. Laboratory/Shop
___5. Resource Room
___6. Other _____

Provide the following information for the class observed.

3. _____ Grade level(s) of students

4. _____ Total number of students in class

5. _____ Number of students having an IEP

6. _____ Intern's number of completed years teaching experience

KDE/MIC App.
3018 1054 3'89

Length of coding (in minutes) _____

Observer's name (print) _____

Observer's signature _____

Intern's signature _____

LESSON COMPONENTS

TEACHER BEHAVIORS

Treatment		1. states/defines lesson component
		2. explains component clearly
		3. demonstrates component
		4. provides guided practice
		5. provides independent practice

TEACHER BEHAVIORS DURING INSTRUCTION

Communication		1. cues students
		2. uses emphasis
		3. uses challenge/task attractions
		4. uses sarcasm/negative affect
Questioning Techniques		1. asks academic questions
		2. asks multiple questions
		3. allows call-outs
		4. does not provide wait time
		5. guides reciter
		6. does not guide reciter
		7. asks procedural questions
		8. asks unrelated questions
Responses		1. acknowledges student responses
		2. rephrases /amplifies student responses
		3. corrects/clarifies student responses
		4. does not correct/clarify responses
Praise		1. uses specific academic praise
		2. uses general academic praise
		3. uses group academic praise

DURING SEATWORK

Management		1. uses specific conduct-related praise
		2. stops misconduct positively
		3. stops misconduct negatively
		4. does not stop misconduct
		5. manages overlapping events
Time		1. minimizes management time
		2. mismanages instructional time

OBSERVATION NOTES:

Lesson Initiation
1._____began observed lesson promptly.
2._____secured the attention of students before beginning instruction.
3._____used a review of students' previous work or previous lesson content to establish academic focus.
4._____provided an over view of lesson content or a statement of lesson objective(s) to establish academic focus.
5._____provided an over view of how the lesson would proceed to establish academic focus.

TEACHER BEHAVIORS TO BE CHECKED AT END OF LESSON
DURING THIS OBSERVATION, THE INTERN:

1. _____ asked questions, paused, then identified respondents.

2. _____ elicited responses from a variety of students in a nonrepetitive pattern to check for understanding or

   elicited responses from students in ordered turns to check for understanding (appropriate only in K-3 small group instruction in math and reading).

3. _____ used appropriate grammar and vocabulary.

4. _____ conducted a review or summary of major lesson components.

5. _____ provided clear directions for seatwork/homework. (Enter NA if not applicable.)

6. _____ circulated and monitored students' progress during seatwork. (Enter NA if not applicable.)

7. _____ was involved.

8. _____ was enthusiastic.

OBSERVER'S SUMMARY STATEMENT (OPTIONAL):

AFTER COMPLETING AN OBSERVATION, PLEASE CHECK TO INSURE THAT:

1. ALL INFORMATION ON THE FRONT PAGE OF THE INSTRUMENT IS COMPLETE
2. EACH LESSON COMPONENT HAS A CHECK (✓) OR ZERO (0) ENTERED FOR EACH TREATMENT BEHAVIOR
3. THE NUMBER OF TALLIES HAS BEEN ENTERED IN THE BOX TO THE LEFT OF EACH BEHAVIOR
4. EACH LESSON INITIATION AND END OF LESSON BEHAVIOR HAS BEEN MARKED WITH A CHECK (✓), ZERO (0), OR NA